N-227
17Apr78

University of Queensland

PRENTICE COMPUTER CENTRE

Newsletter

authorization: Director of the Centre

1 CONTENTS OF THIS ISSUE

2 SUNDAY OPERATION

Sunday operation was introduced during peak load periods when sufficient capacity was not available during normal hours of operation. The charges are half the normal charge rate.

The new PDP1090 system will be brought into production during the weekend of 22-23 April. As you can appreciate, we are anxious to keep our costs and hence our charges to a minimum. For this reason, we have not recruited a full establishment of operators for the new system. We are relying on the fact that the additional capacity now available will enable the Centre to cease third shift and weekend operation for a period.

We believe that there will be sufficient capacity to meet users requirements at acceptable response levels.

I accept that there may be some users who, for other reasons, would prefer to work on a Sunday. However, it is difficult to justify the cost of operation for a relatively small number.

Sunday working will cease after 16 April. I suggest that we wait and see if the implementation of the new 1090 system provides relief and allows users to schedule their time with greater effectiveness than was previously the case.

If it is thought by users that Sunday operation should be re-introduced, I would be grateful to receive some firm proposals to allow me to make a case for the additional resources needed.

Director (2189)

3 CORRECTIONS TO THE EDITOR PROGRAM, QEDIT

The editor has been on the system without change for six months. Since it is the most extensively used program on the system, usually five persons using it at any time, one would have thought that any errors in it would have been observed by now. Although no-one reported it, the string form of DINSERT did not work. This has been corrected and other changes made to facilitate editing of very long lines.

Very long lines are split on input into 280 character chunks (formerly 300 characters) leaving room for changes to add up to 20 more characters. Each chunk is treated during editing as if it was a line, but they are concatenated into lines again on output.

Please report any editing difficulties or apparent errors to the consulting programmer. He will refer genuine errors to me for prompt attention.

Ian Burgess (3944)

4 EFFICIENCY HINTS WITH SPSS

The following is reprinted from the SPSS newsletter produced by the originators of the package:

Although we cannot be much help in predicting the precise cost of your job, we can list a few general rules which if followed will usually make your runs cheaper.

1. ASK ONLY FOR THOSE STATISTICS YOU REALLY WANT. Some statistics have relatively little overhead, others can be quite expensive to compute. By and large, the defaults which were chosen are relatively inexpensive to compute.

2. USE ONLY THOSE FILES THAT YOU NEED. SPSS writes a scratch file during the first procedure which is then used for subsequent procedures and for SAVE FILE. If you have only one procedure in your run and are not going to save a file, then dummy the scratch file.

On the DEC-10, this can be achieved by use of the /SCRATCH switch, thus

```
.R STA:SPSS
*myprog.sps,NUL:/SCRATCH
```

Some procedures do require a second pass of the data depending upon options and statistics requested. The scratch file may not be dummied for Z scores in CONDESCRIPTIVE (Option 3); REGRESSION with keyword RESID; FACTOR with keyword FACSCORE; or CANCORR with kayword CANVAR. Nor may the scratch file be dummied for the classification options in DISCRIMINANT (Options 5, 6, 7 and 8). The scratch file may not be dummied if subfiles are processed out of order.

3. TAKE ADVANTAGE OF MISSING VALUES OPTIONS. If no missing values are defined for the variables being analyzed in a procedure, then select Option 1 (inclusion of missing data) to avoid having the system spend time checking for missing values. Inclusion of missing values of LISTWISE deletion is always faster and requires less core in the procedures based on correlation techniques than does PAIRWISE deletion.

4. DO MULTIPLE PROCEDURES DURING A RUN. SPSS writes a scratch file for the first procedure and accesses it for each subsequent procedure. As the scratch file contains all permanent data transformations and data selections, its repeated use in runs with several procedures represents a considerable efficiency. Deformatting raw BCD takes time.

5. INCREASE WORKSPACE INSTEAD OF SPLITTING UP PROCEDURE REQUESTS. A procedure card requesting many options and statistics may require a large amount of WORKSPACE, but it is usually cheaper to request extra WORKSPACE than to process multiple procedure requests requiring less WORKSPACE as each procedure card encountered forces the data to be passed.

6. CHOOSE YOUR PROCEDURES CAREFULLY. The execution times for integer FREQUENCIES, CROSSTABS, and BREAKDOWN are substantially shorter than those for the general versions of those procedures. However, the factors relevant to determining processing times for integer and general versions are different. The most important factor for the general versions is the product of the total number of cells requested and the number of cases; for the integer versions, it is the product of the number of tables requested and the number of cases. Users should also note that if the matrix of cells is sparse, the excessive amounts of core required by integer versions may offset the efficiencies of their collecting algorithms, making a general version a better choice in some instances.

Note also that if the file is sorted correctly, AGGREGATE may often be used in place of BREAKDOWN, and is much less expensive.

7. CHOOSE YOUR DATA TRANSFORMATIONS CAREFULLY. RECODE's and COUNT's are the cheapest, IF's the most expensive.

8. SAMPLE! SAMPLE can be used, especially during exploratory analysis, to cut costs.

9. PUNCH AND SAVE CORRELATION MATRICES. Much of the cost of REGRESSION, FACTOR, and PARTIAL CORR is in the calculation of the correlation matrix. If the matrix can be reused, you will save money.

   To direct to which file you wish the matrix to be written use the RAW OUTPUT UNIT card to specify a filename, directly before your procedure card,

```
example -        RAW OUTPUT UNIT    MATRIX.DAT
                 REGRESSION    ------
                      .
                      .
                      .
```

   Of course, input of the correlation matrix closes off the users options of obtaining factor scores etc. which require recourse to the raw data as well.

10. SAVE AND USE SYSTEM FILES. They are self-documenting and reflect all your "data massaging".

Chris McGovern (3968)

5 ERRORS CORRECTED IN SPSS

The following major errors in version 7 of SPSS, which was installed last month, have been corrected.

(1) MULT RESPONSE

Runs aborted with an "Illegal Instruction @PC...."

(2) FREQUENCIES

General mode gave "?Ill mem ref" when tabulating a variable with a large (>342) number of distinct values.

(3) SAVE FILE

An infinite loop could result if an error was encountered in the deck before pass one over the data was done.

Chris McGovern (3968)

* * * * *